Derivative-free optimization of composite functions

Jeffrey Larson

Argonne National Laboratory

May 27, 2019

Problem setup

$$\underset{x}{\text{minimize}}\ f(x) = h(x; S(x))$$
$$\text{subject to: } x \in \mathcal{D} \subset \mathbb{R}^n$$

where the objective f depends on the output(s) from a simulation S and a known function h.

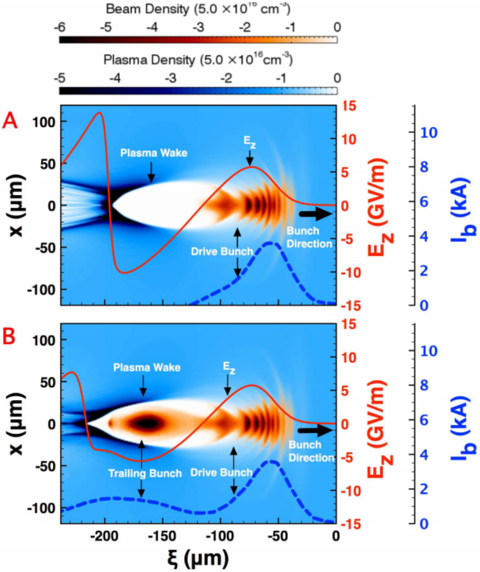
Problem setup

$$\underset{x}{\text{minimize}}\ f(x) = h(x; S(x))$$
$$\text{subject to: } x \in \mathcal{D} \subset \mathbb{R}^n$$

where the objective f depends on the output(s) from a simulation S and a known function h.

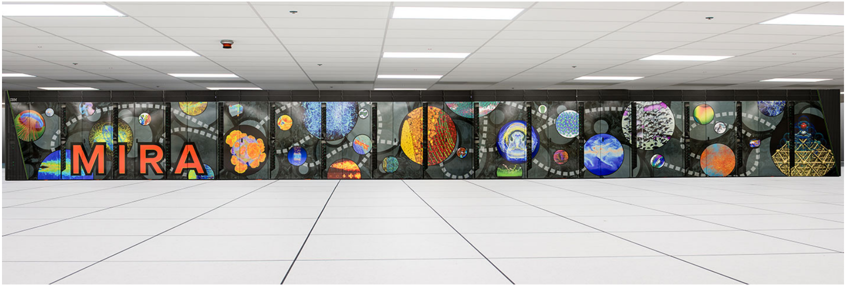
- Derivatives of S may not be available
- Constraints defining $\mathcal{D}$ may or may not depend on S
- The dimension n is small
- Evaluating S is expensive (not using grids or randomized/evolutionary methods)

Computers/Simulations!



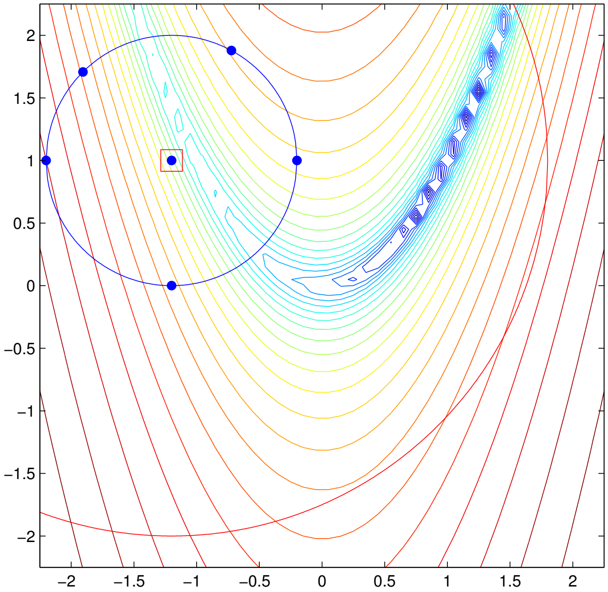
Computers/Simulations!

Computers/Simulations!

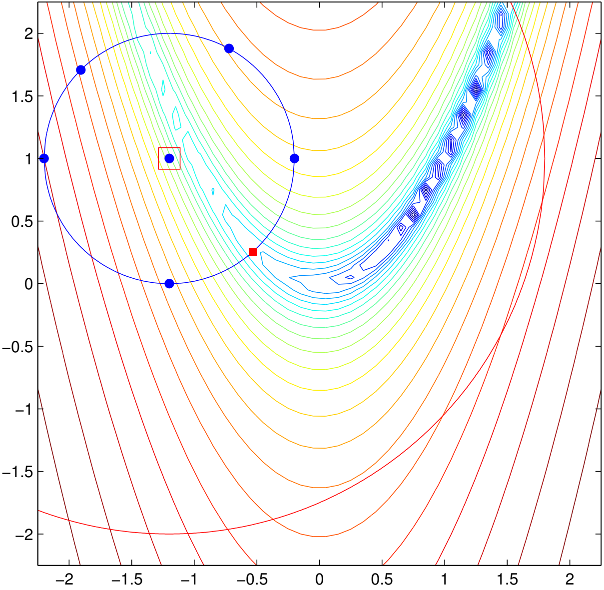


#21 on TOP500 November 2018 (#6 on TOP500 June 2016)

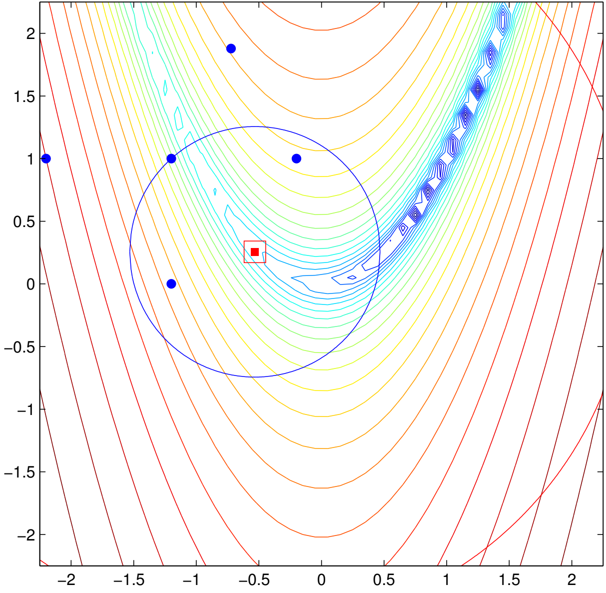
Model-based methods - Interpolation



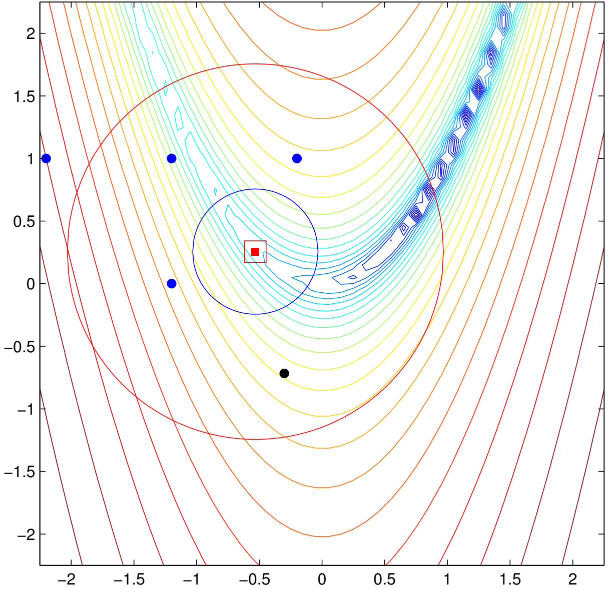
Model-based methods - Interpolation



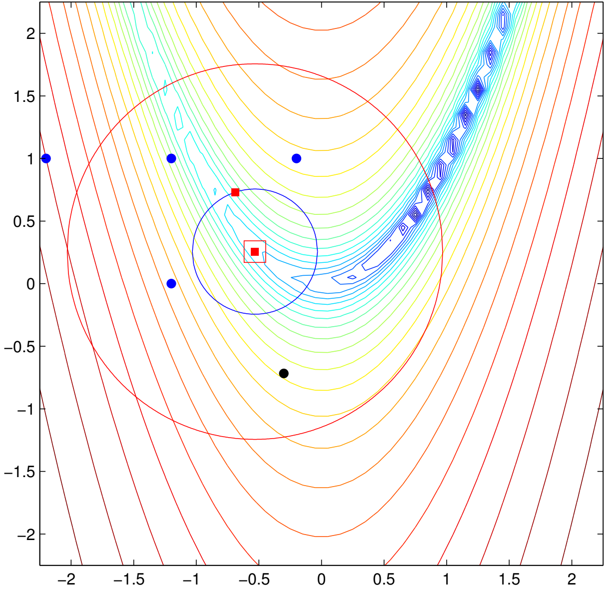
Model-based methods - Interpolation



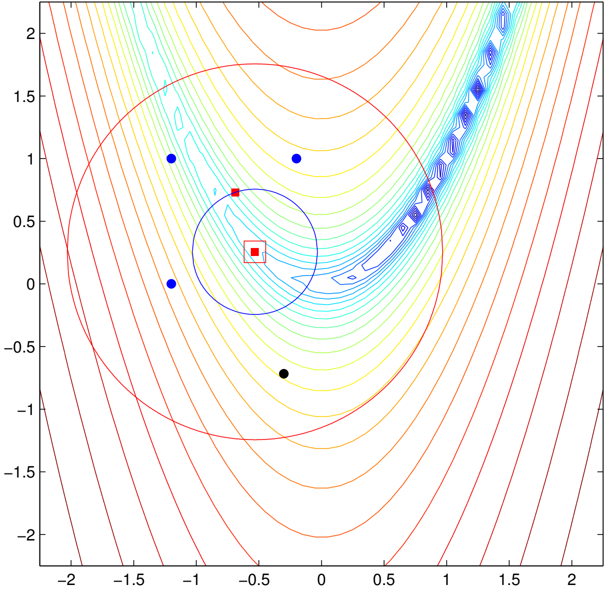
Model-based methods - Interpolation



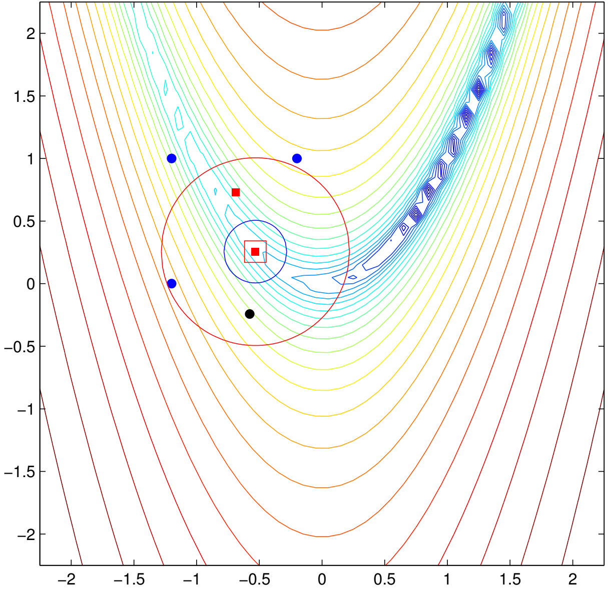
Model-based methods - Interpolation



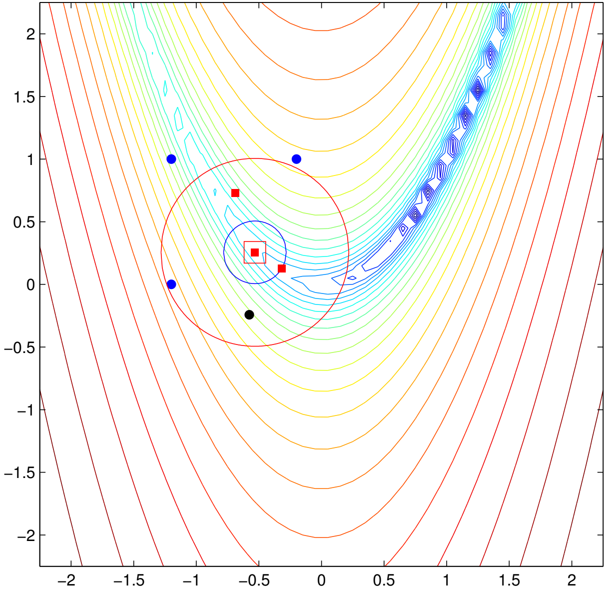
Model-based methods - Interpolation



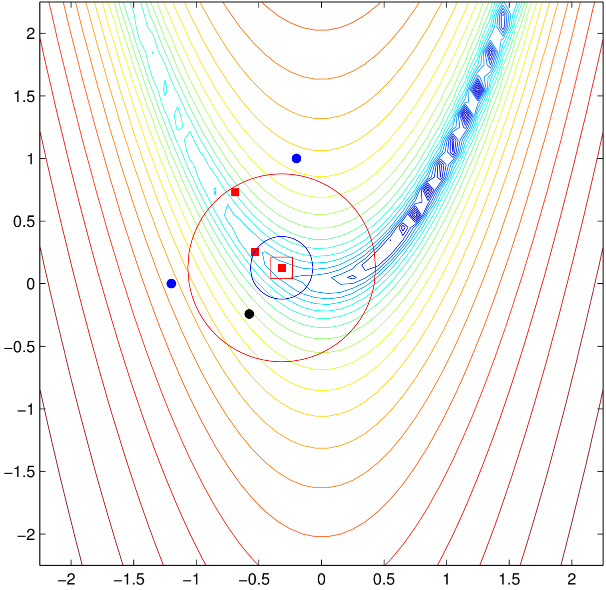
Model-based methods - Interpolation



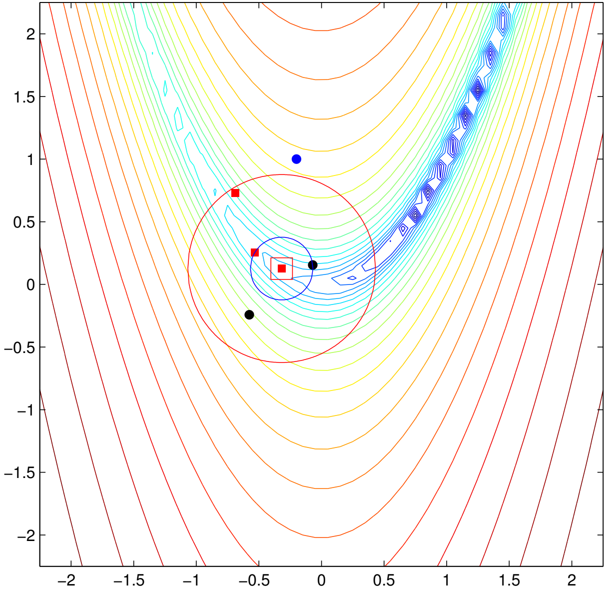
Model-based methods - Interpolation



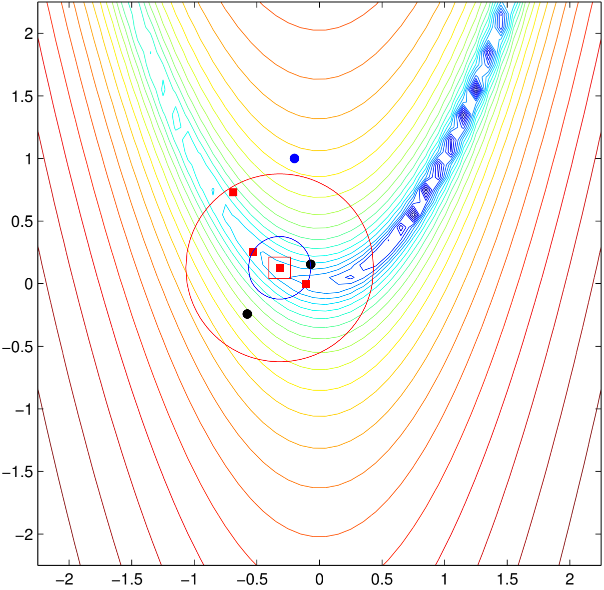
Model-based methods - Interpolation



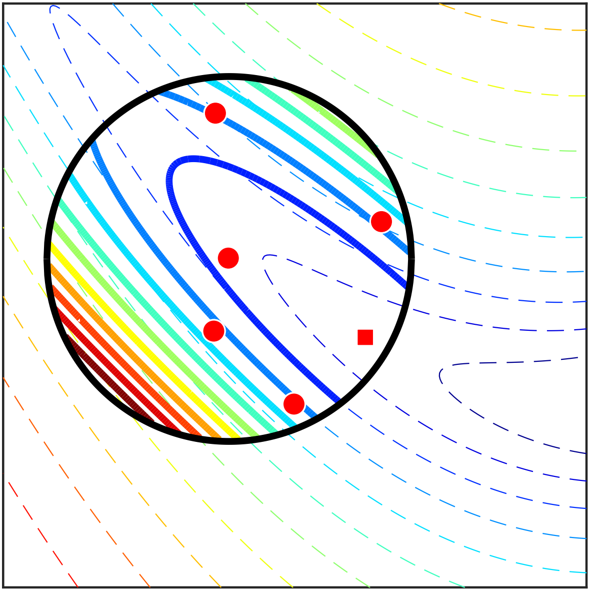
Model-based methods - Interpolation



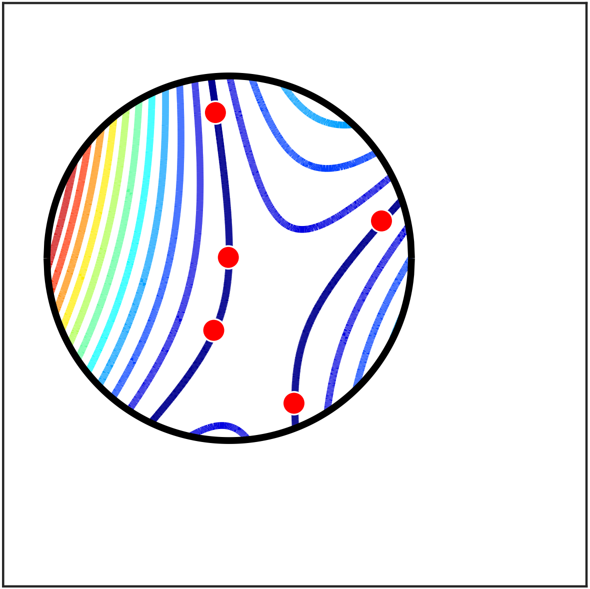
Model-based methods - Interpolation



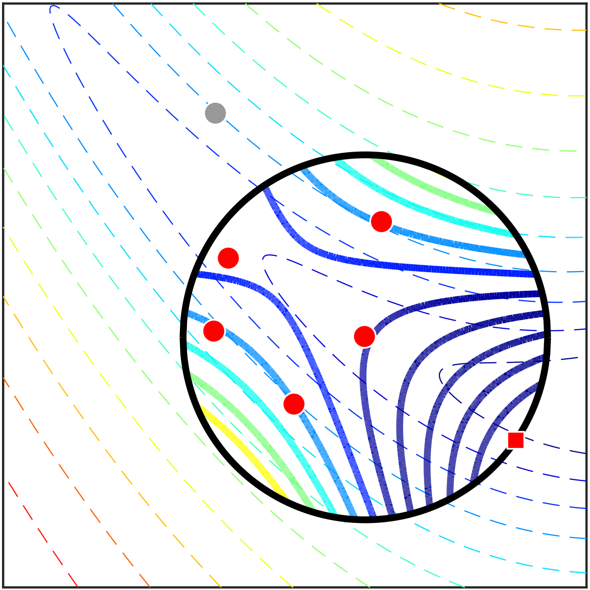
Model-based methods - Interpolation



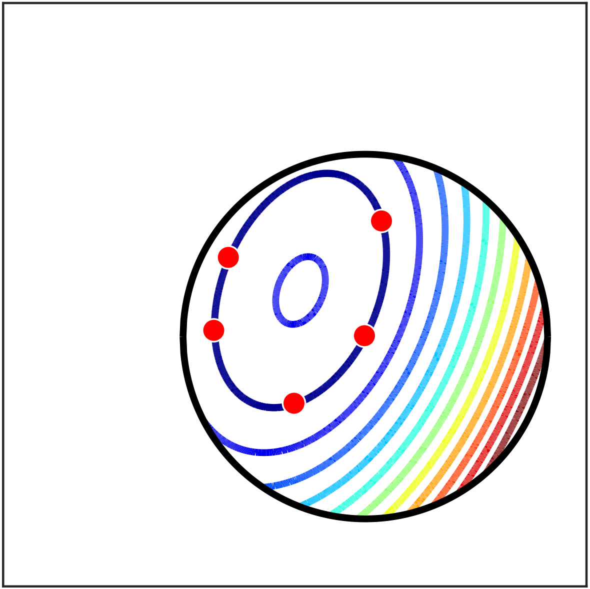
Model-based methods - Interpolation



Model-based methods - Interpolation



Model-based methods - Interpolation

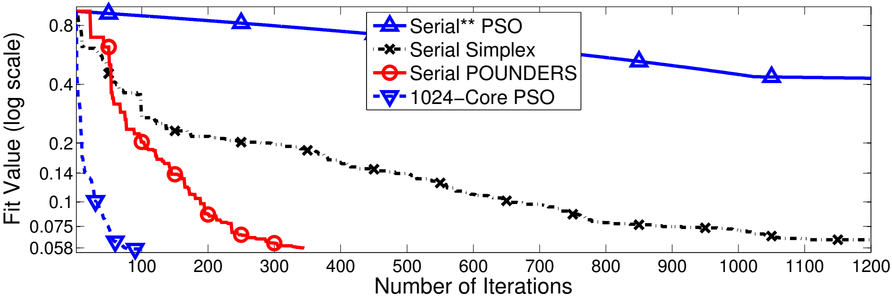


Opening up the black box

$$f(x) = \|S(x) - T\|_2^2 = \sum_{i=1}^{p}(S_i(x) - T_i)^2$$

Can either have a solver that uses $f(x)$ or $[S_1(x), \ldots, S_p(x)]$.

Opening up the black box



Tuning quadrupole moments for a particle accelerator simulation.

$$f(x) = \|S(x) - T\|_2^2 = \sum_{i=1}^{p} (S_i(x) - T_i)^2$$

Can either have a solver that uses $f(x)$ or $[S_1(x), \ldots, S_p(x)]$.

Emittance minimization

$$\underset{v \in \mathbb{R}^n}{\text{minimize}}\ \epsilon(v)$$

$$\textbf{subject to: } v \in \mathcal{D}$$

where

$$\epsilon(v) = \sqrt{\langle x(v)^2 \rangle \langle p_x(v)^2 \rangle - \langle x p_x(v) \rangle^2}$$

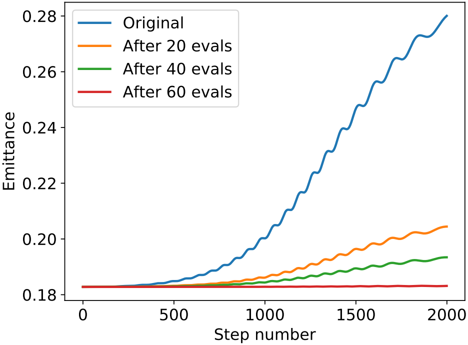
Emittance minimization

$$\underset{v \in \mathbb{R}^n}{\text{minimize}}\ \epsilon(v)$$

$$\text{subject to: } v \in \mathcal{D}$$

where

$$\epsilon(v) = \sqrt{\langle x(v)^2 \rangle \langle p_x(v)^2 \rangle - \langle x p_x(v) \rangle^2}$$



Emittance minimization

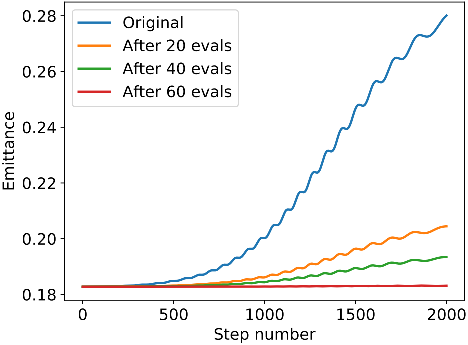
$$\underset{v \in \mathbb{R}^n}{\text{minimize}} \ \min_t \epsilon(v, t)$$

$$\textbf{subject to: } v \in \mathcal{D}$$

$$0 \leq t \leq \bar{t},$$

where

$$\epsilon(v, t) = \sqrt{\langle x(v, t)^2 \rangle \langle p_x(v, t)^2 \rangle - \langle x p_x(v, t) \rangle^2}$$



Exploiting Structure

- Nonsmooth, composite optimization

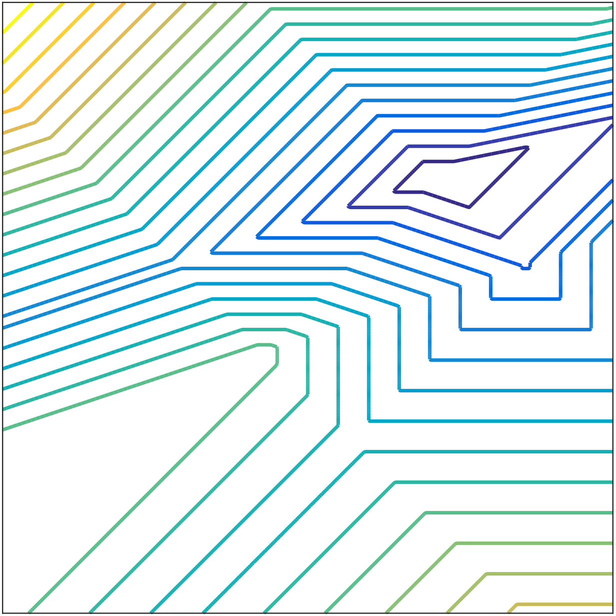
$$\underset{x}{\text{minimize}}\, f(x) = h(S(x))$$

nonsmooth $h\colon \mathbb{R}^p \to \mathbb{R}$ (with a known structure), smooth $S\colon \mathbb{R}^n \to \mathbb{R}^p$ (expensive to evaluate).

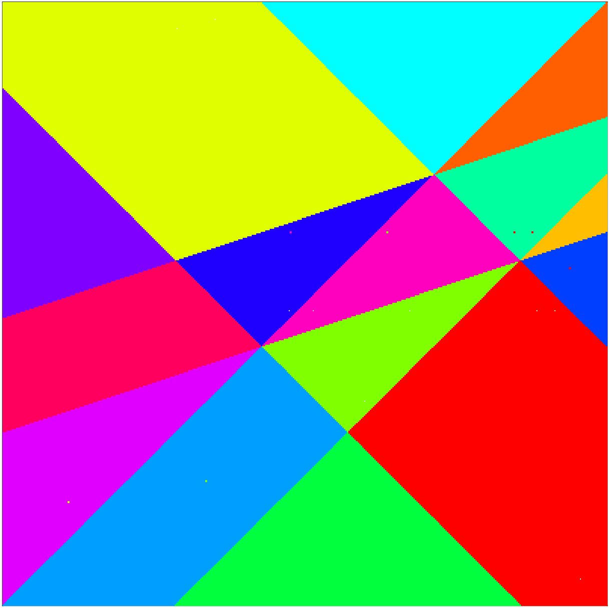
- Idea: Build p models, one for each component of S. Use model gradients in place of ∇S.
- Requires a *manifold representation* of h.
- Example: censored ℓ_1 loss:

$$f(x) = \sum_{i=1}^{p} |d_i - \max\{c_i, S_i(x)\}|$$

Censored ℓ_1 loss



Censored ℓ_1 loss



Manifold representation

- $h(y) = \max_{i \in \{1,\ldots,p\}} y_i$ $\qquad$ p **manifolds**

Manifold representation

- $h(y) = \max_{i \in \{1,\ldots,p\}} y_i$ $\quad$ p **manifolds**

- $h(y) = \|y\|_\infty = \max_{i \in \{1,\ldots,p\}} |y_i|$ $\quad$ $2p$ **manifolds**

Manifold representation

- $h(y) = \max_{i \in \{1,\ldots,p\}} y_i$ $\quad$ p **manifolds**

- $h(y) = \|y\|_\infty = \max_{i \in \{1,\ldots,p\}} |y_i|$ $\quad$ $2p$ **manifolds**

- $h(y) = \|y\|_1 = \sum_{i=1}^{p} |y_i|$ $\quad$ 2^p **manifolds**

Manifold representation

- $h(y) = \max_{i \in \{1,\dots,p\}} y_i$ $\qquad$ p **manifolds**

- $h(y) = \|y\|_\infty = \max_{i \in \{1,\dots,p\}} |y_i|$ $\qquad$ $2p$ **manifolds**

- $h(y) = \|y\|_1 = \sum_{i=1}^{p} |y_i|$ $\qquad$ 2^p **manifolds**

- $h(y) = \sum_{i=1}^{p} |d_i - \max\{c_i, y_i\}|$ $\qquad$ 3^p **manifolds. If** $p = 45$**, approximately** 3×10^{21} **potential manifolds.**

Manifold representation

- $h(y) = \max_{i \in \{1,\ldots,p\}} y_i$ $\quad$ p **manifolds**

- $h(y) = \|y\|_\infty = \max_{i \in \{1,\ldots,p\}} |y_i|$ $\quad$ $2p$ **manifolds**

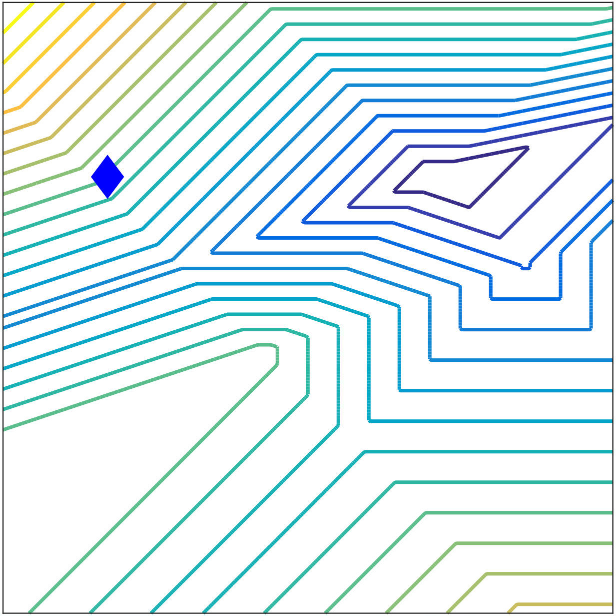
- $h(y) = \|y\|_1 = \sum_{i=1}^{p} |y_i|$ $\quad$ 2^p **manifolds**

- $h(y) = \sum_{i=1}^{p} |d_i - \max\{c_i, y_i\}|$ $\quad$ 3^p **manifolds. If** $p = 45$**, approximately** 3×10^{21} **potential manifolds.**

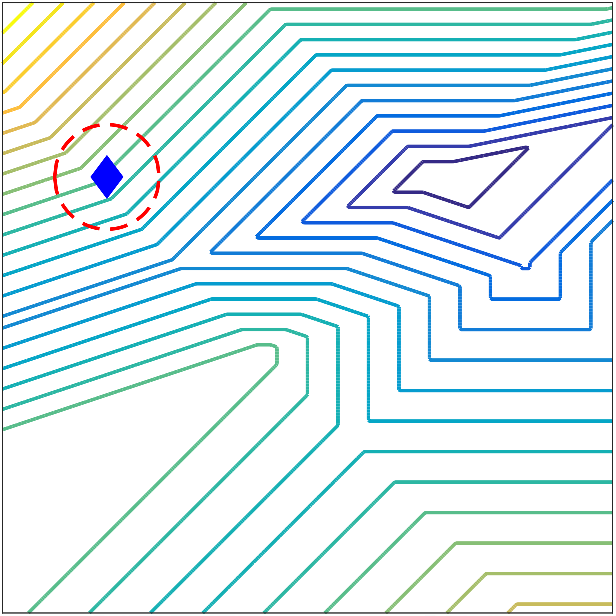
User scripts need to calculate:

$f(x)$, $S(x)$, $\mathbb{H}(S(x))$, $\{\nabla h_i(S(x)) : i \in \mathbb{H}(S(x))\}$, $\{h_i(S(x)) : i \in \mathbb{G}\}$,

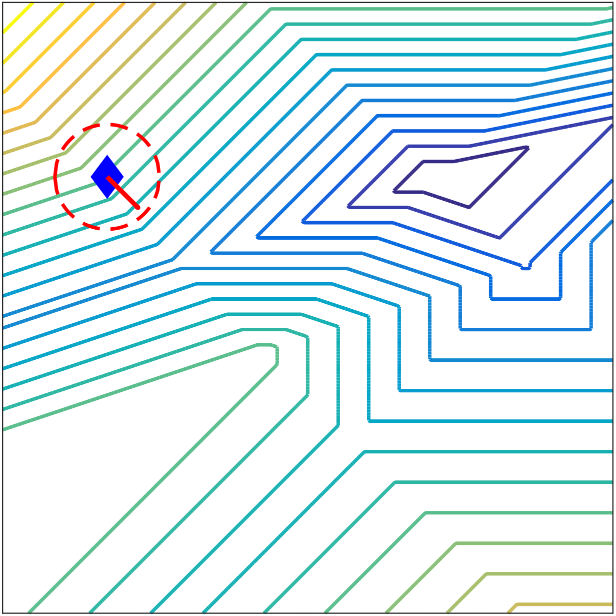
Manifold Sampling



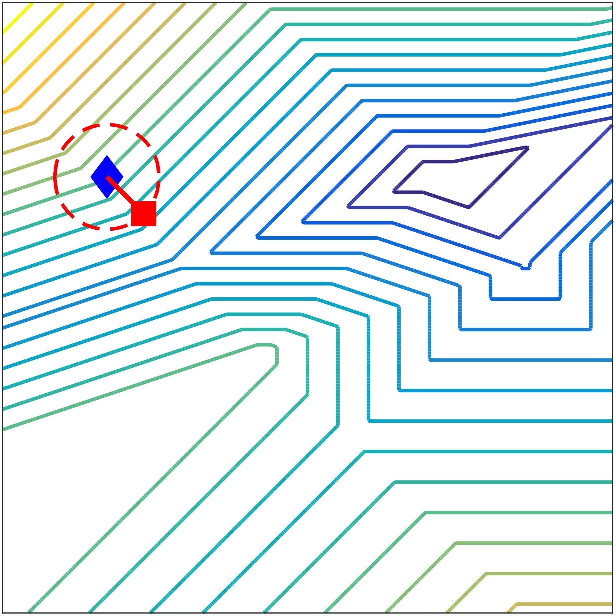
Manifold Sampling



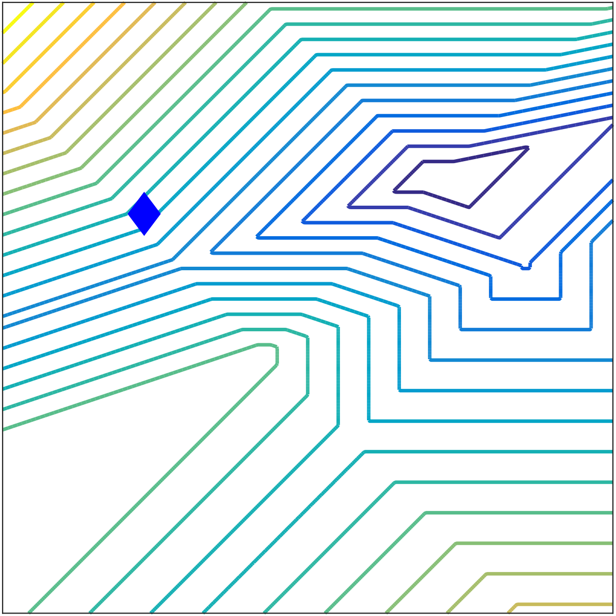
Manifold Sampling



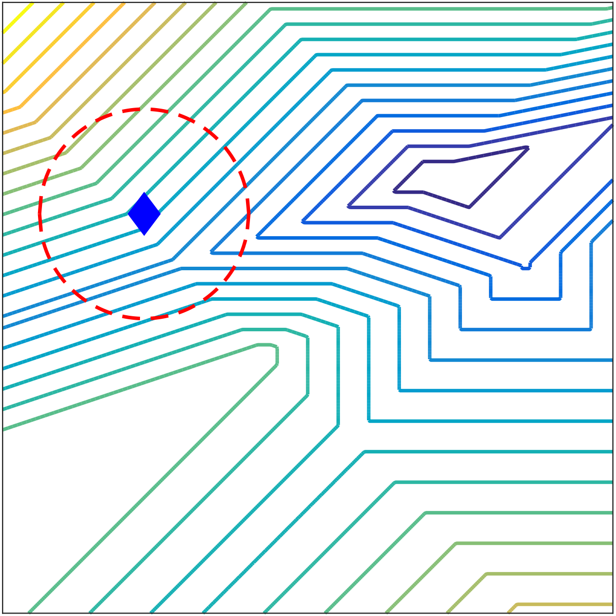
Manifold Sampling



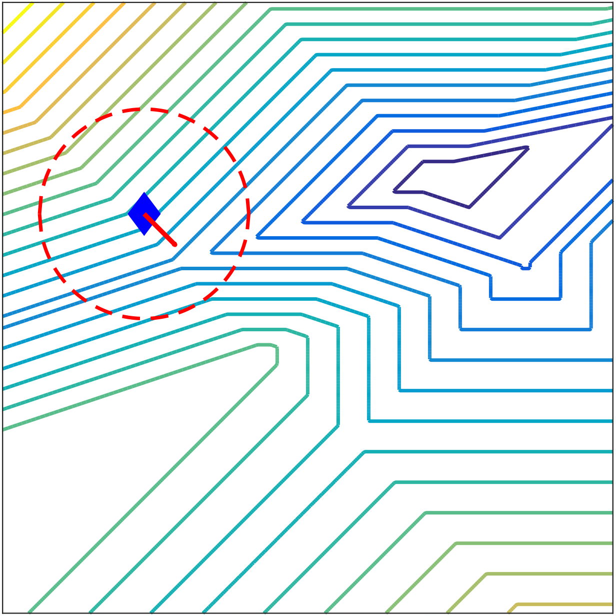
Manifold Sampling



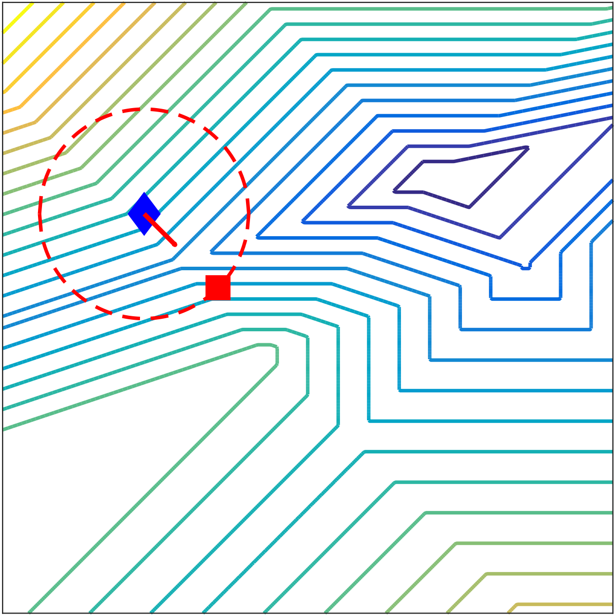
Manifold Sampling



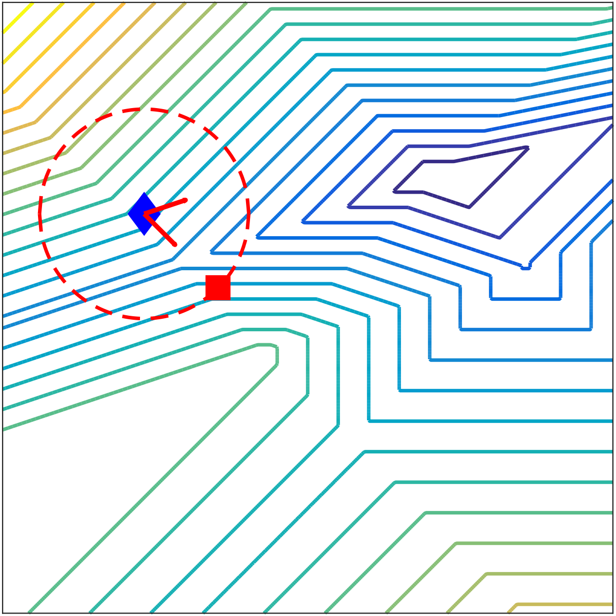
Manifold Sampling



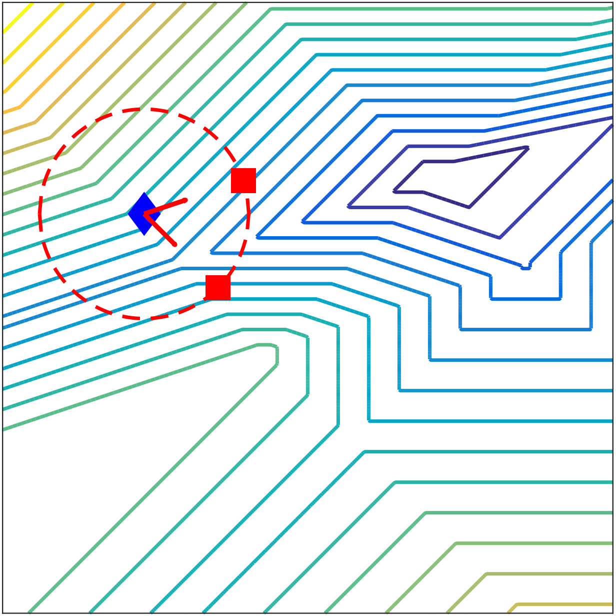
Manifold Sampling



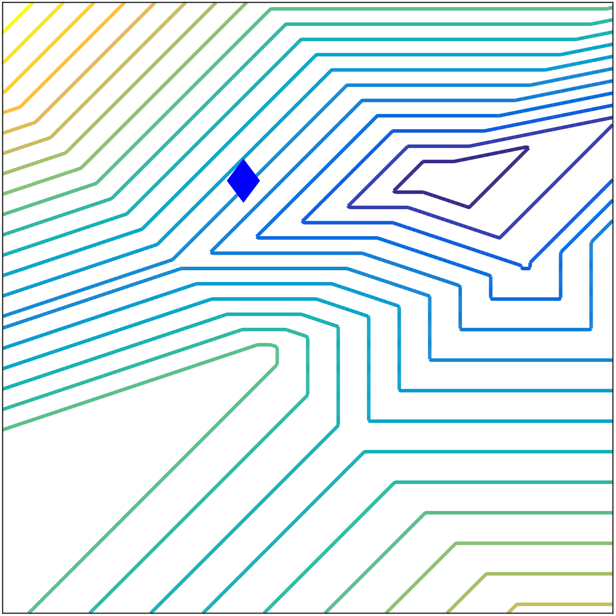
Manifold Sampling



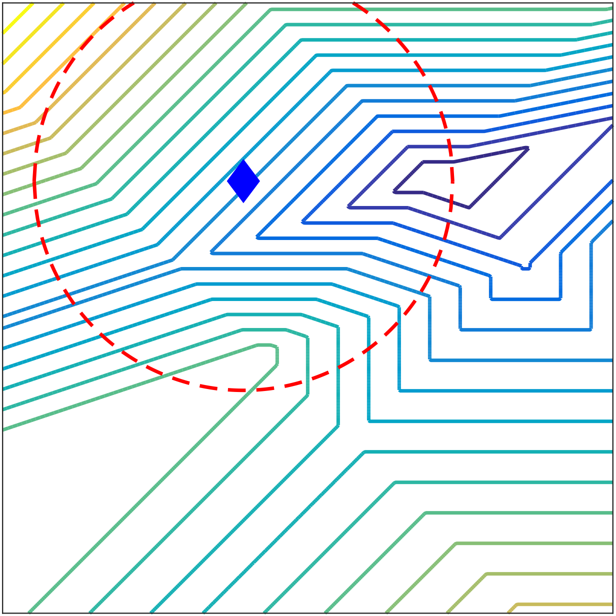
Manifold Sampling



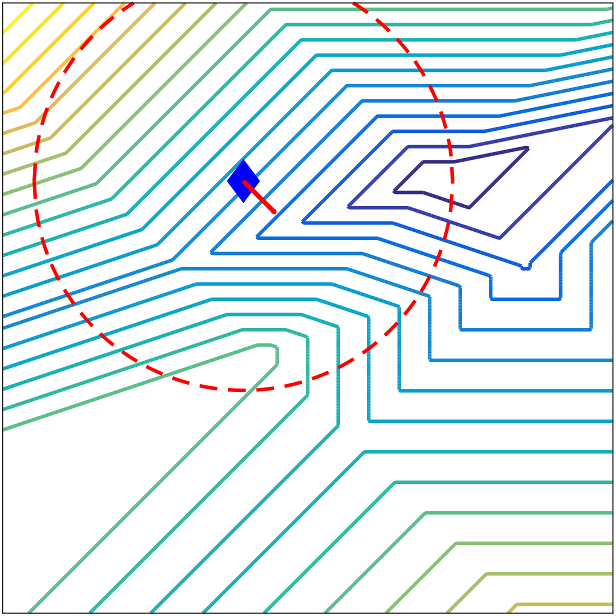
Manifold Sampling



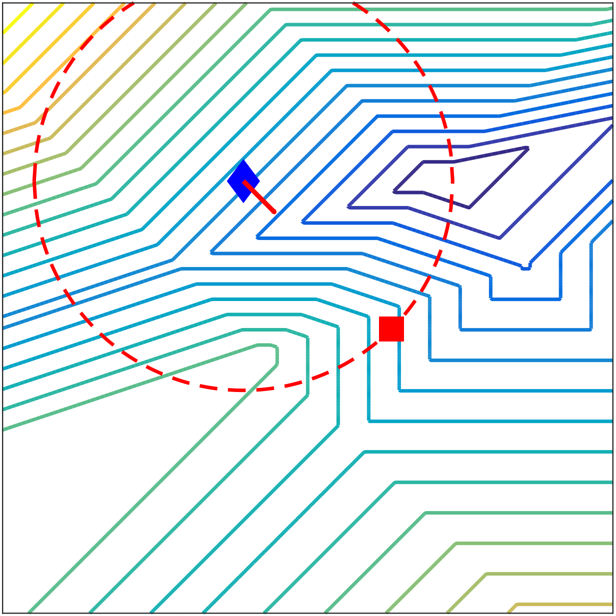
Manifold Sampling



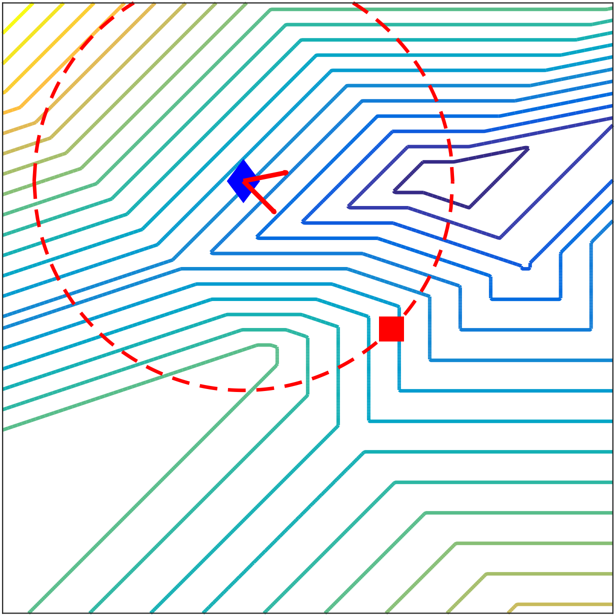
Manifold Sampling



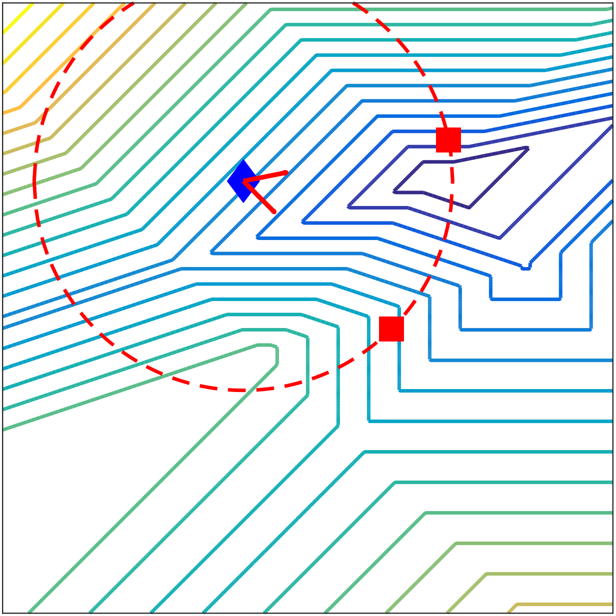
Manifold Sampling



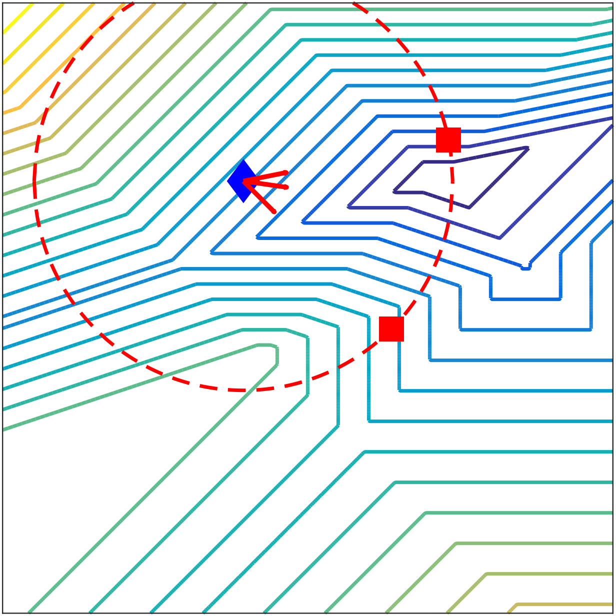
Manifold Sampling



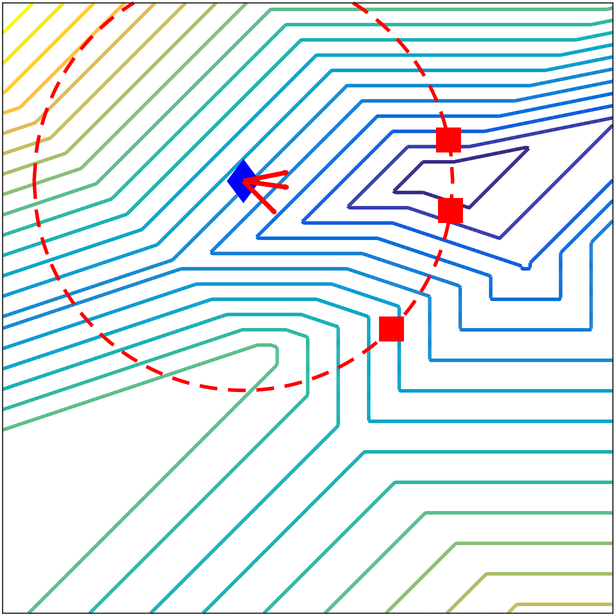
Manifold Sampling



Manifold Sampling



Manifold Sampling



Smooth master model

$$g^k \triangleq \mathbf{proj}\left(0, \mathbf{co}\left(\mathbb{G}^k\right)\right) \in \mathbf{co}\left(\mathbb{G}^k\right),$$

Smooth master model

$$g^k \triangleq \mathbf{proj}\left(0, \mathbf{co}\left(\mathbb{G}^k\right)\right) \in \mathbf{co}\left(\mathbb{G}^k\right),$$

where

$$\mathbb{G}^k \triangleq \bigcup_{i \in I_h(S(x^k))} \left\{\nabla M(x^k) \nabla h_i(S(x^k))\right\}$$

Smooth master model

$$g^k \triangleq \mathbf{proj}\left(0, \mathbf{co}\left(\mathbb{G}^k\right)\right) \in \mathbf{co}\left(\mathbb{G}^k\right),$$

where

$$\mathbb{G}^k \triangleq \bigcup_{i \in I_h(S(x^k))} \left\{\nabla M(x^k) \nabla h_i(S(x^k))\right\}$$

or

$$\mathbb{G}^k \triangleq \bigcup_{y \in Y} \bigcup_{i \in I_h(S(y))} \left\{\nabla M(x^k) \nabla h_i(S(x^k))\right\}$$

Smooth master model

$$g^k \triangleq \textbf{proj}\left(0, \textbf{co}\left(\mathbb{G}^k\right)\right) \in \textbf{co}\left(\mathbb{G}^k\right),$$

where

$$\mathbb{G}^k \triangleq \bigcup_{i \in I_h(S(x^k))} \left\{\nabla M(x^k) \nabla h_i(S(x^k))\right\}$$

or

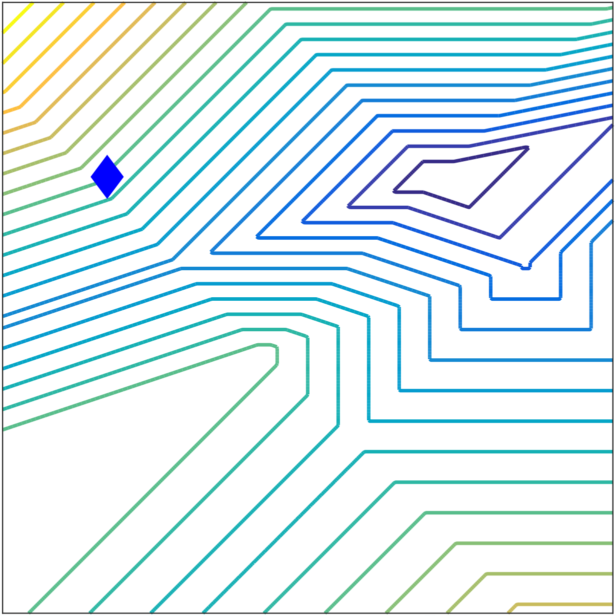
$$\mathbb{G}^k \triangleq \bigcup_{y \in Y} \bigcup_{i \in I_h(S(y))} \left\{\nabla M(x^k) \nabla h_i(S(x^k))\right\}$$

Define the smooth *master model* $m_k^f : \mathbb{R}^n \to \mathbb{R}$ (with gradient g^k) and obtain step by (approximately) solving

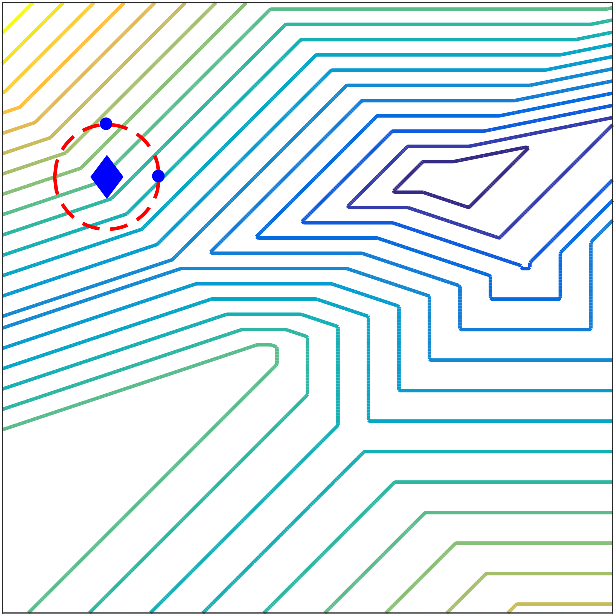
$$\underset{s}{\text{minimize}}\; m_k^f(x^k + s)$$

$$\text{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

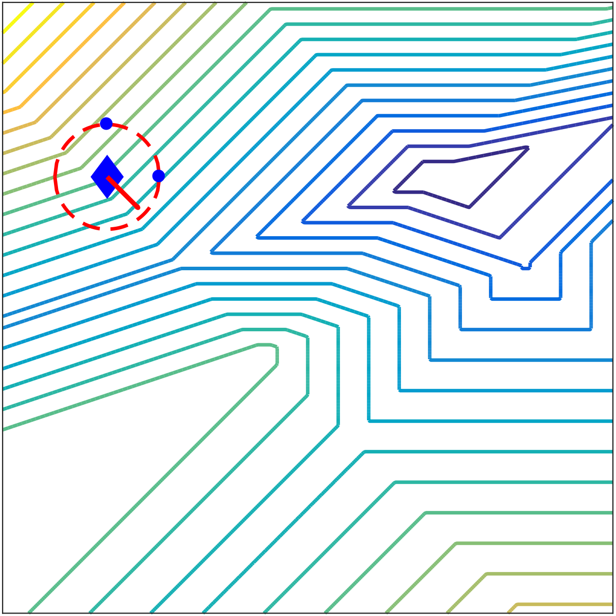
Manifold Sampling



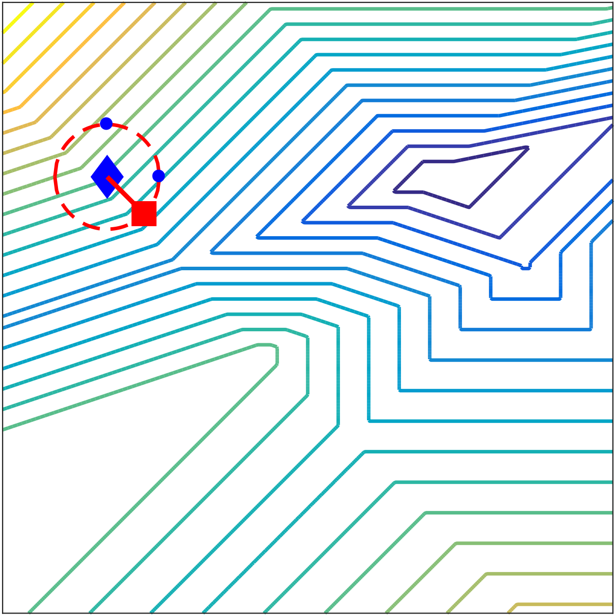
Manifold Sampling



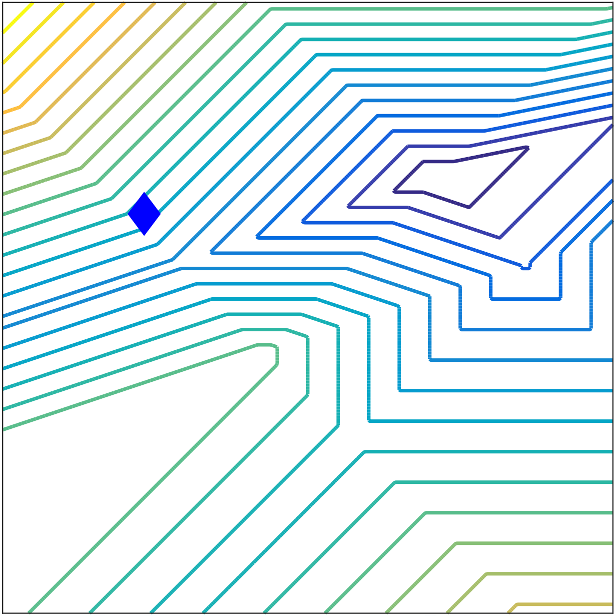
Manifold Sampling



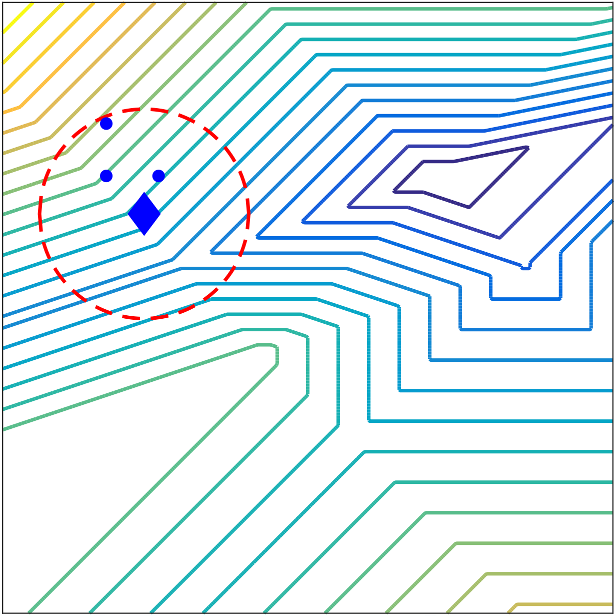
Manifold Sampling



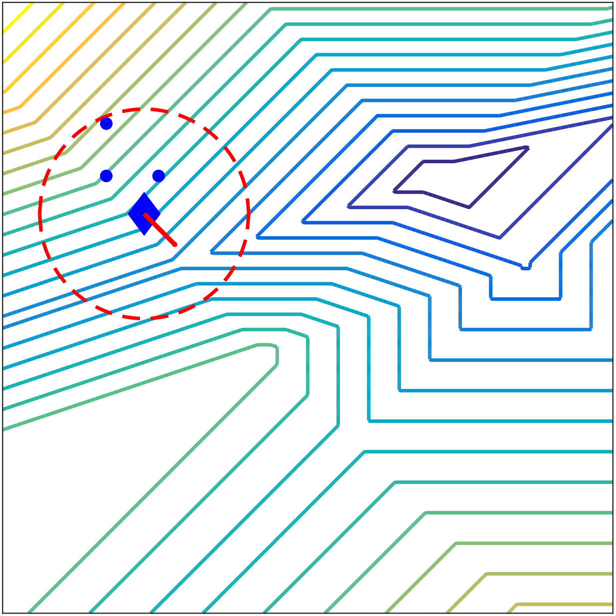
Manifold Sampling



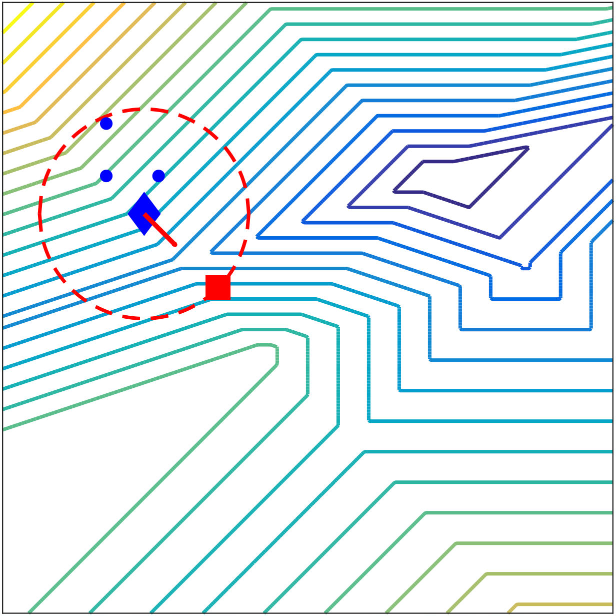
Manifold Sampling



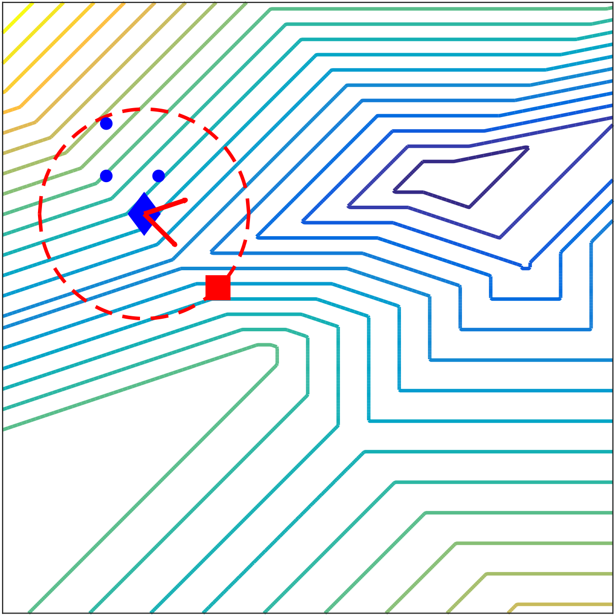
Manifold Sampling



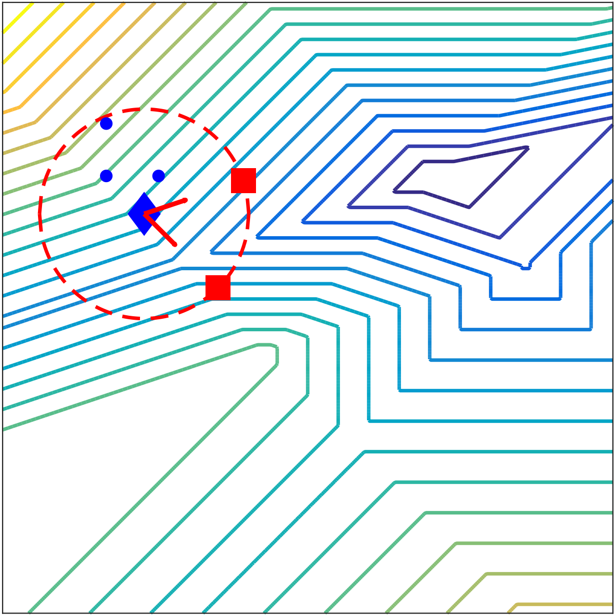
Manifold Sampling



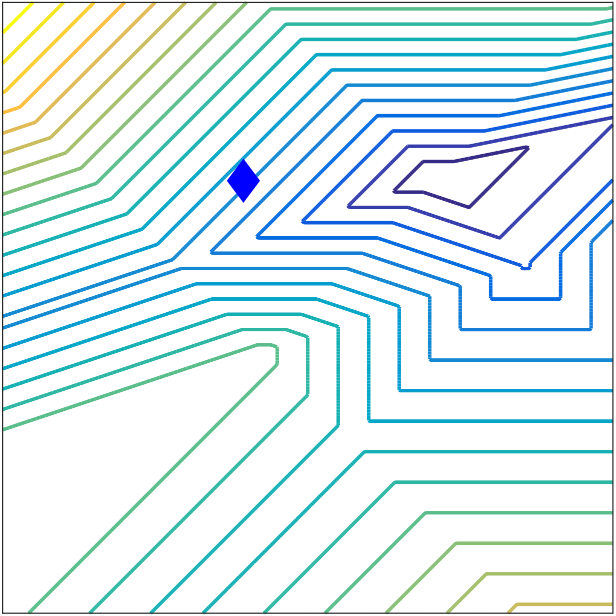
Manifold Sampling



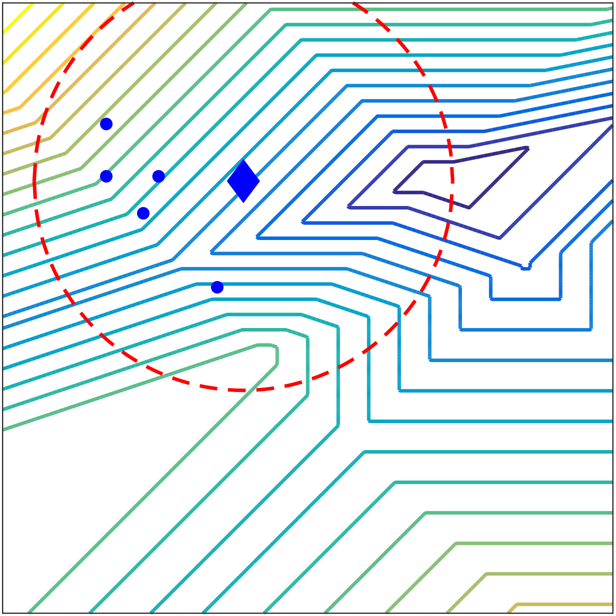
Manifold Sampling



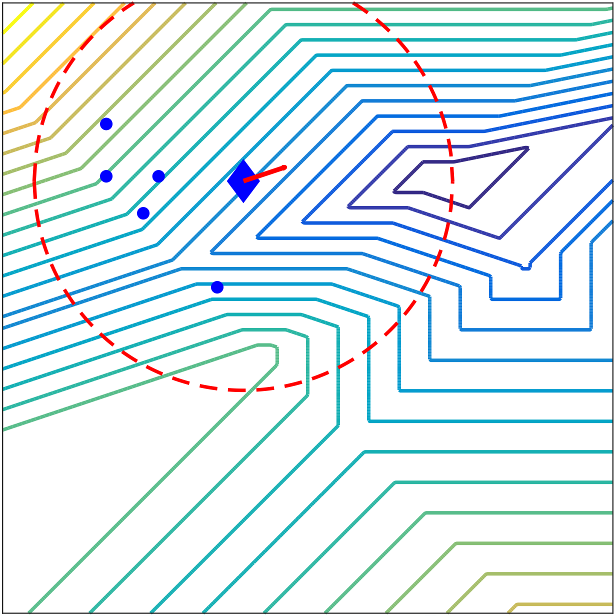
Manifold Sampling



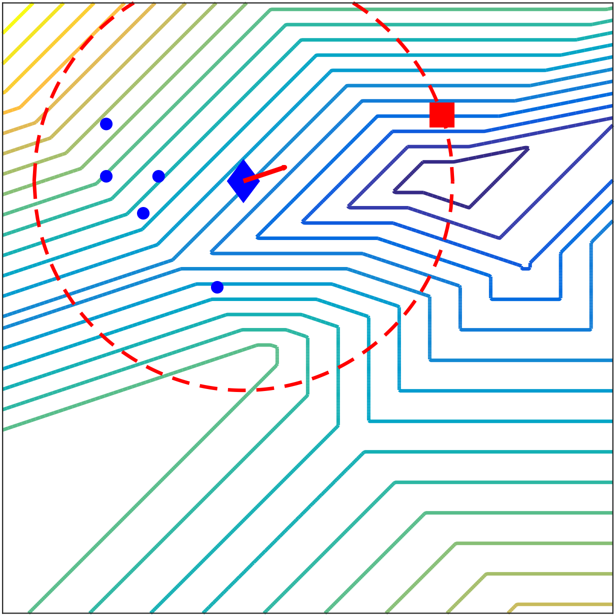
Manifold Sampling



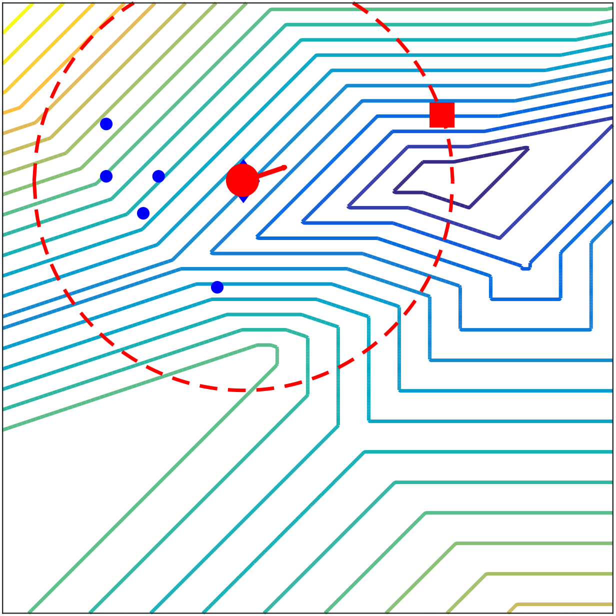
Manifold Sampling



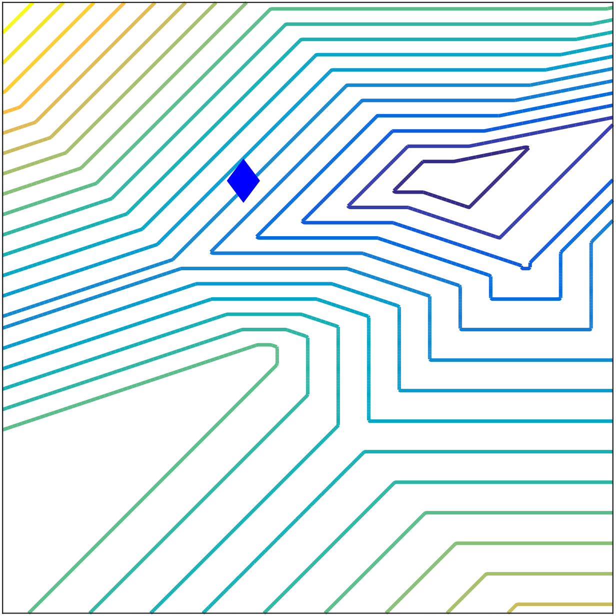
Manifold Sampling



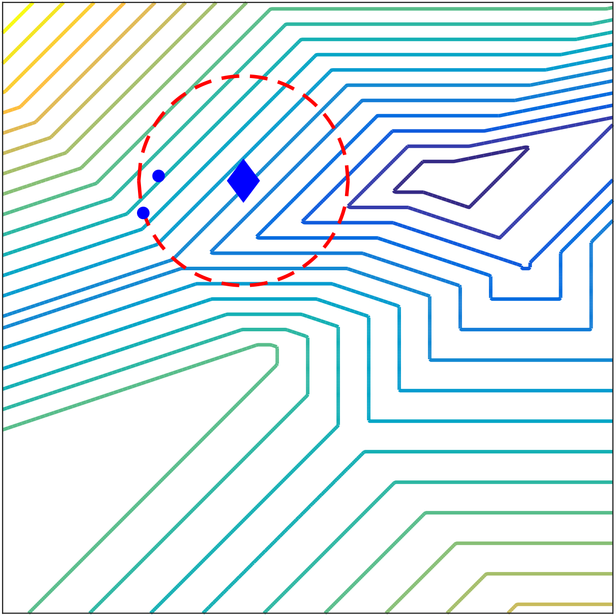
Manifold Sampling



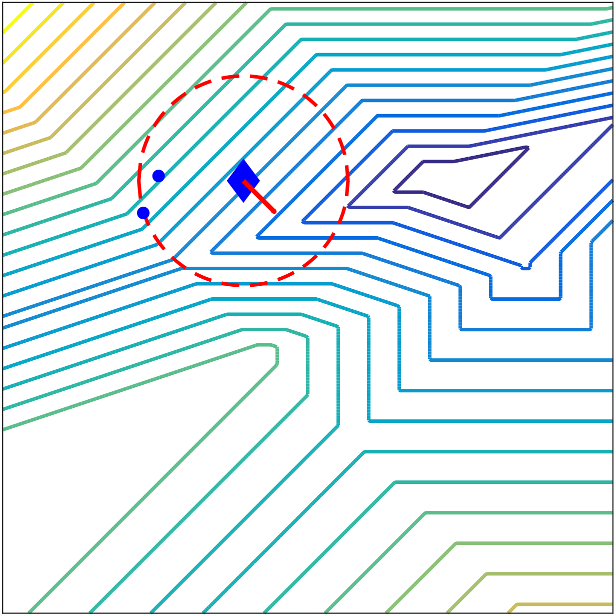
Manifold Sampling



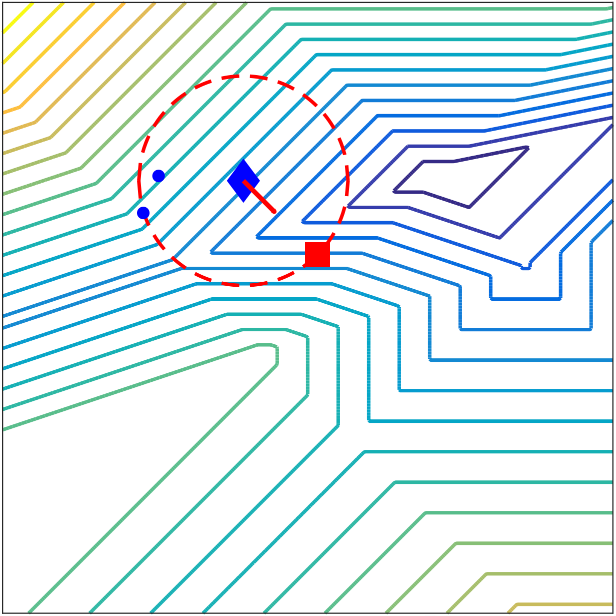
Manifold Sampling



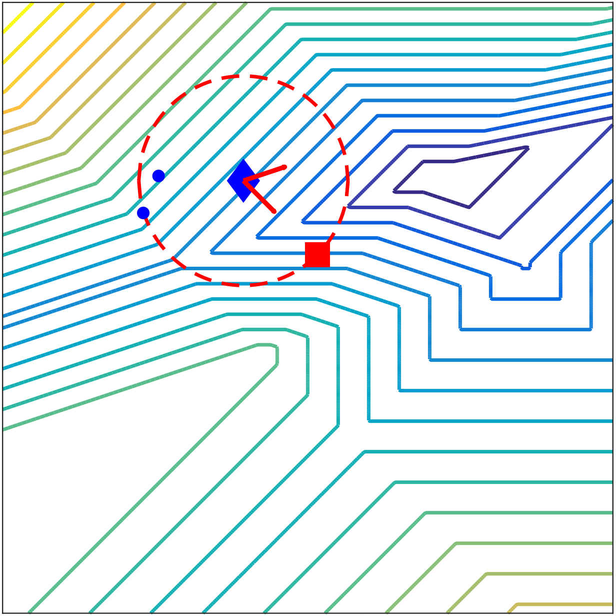
Manifold Sampling



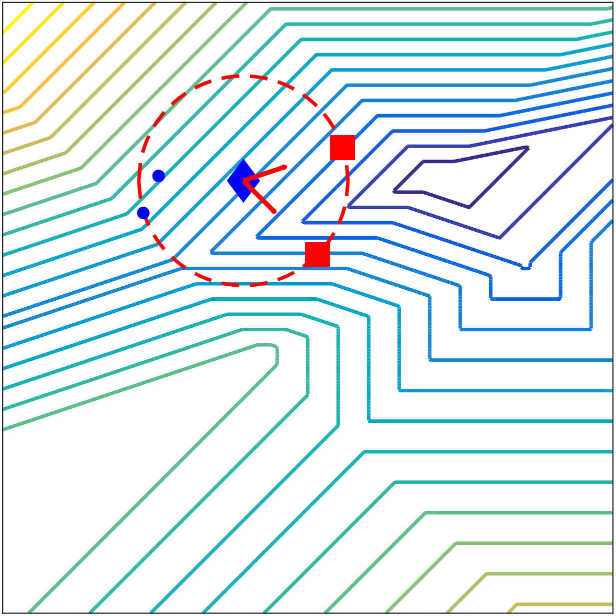
Manifold Sampling



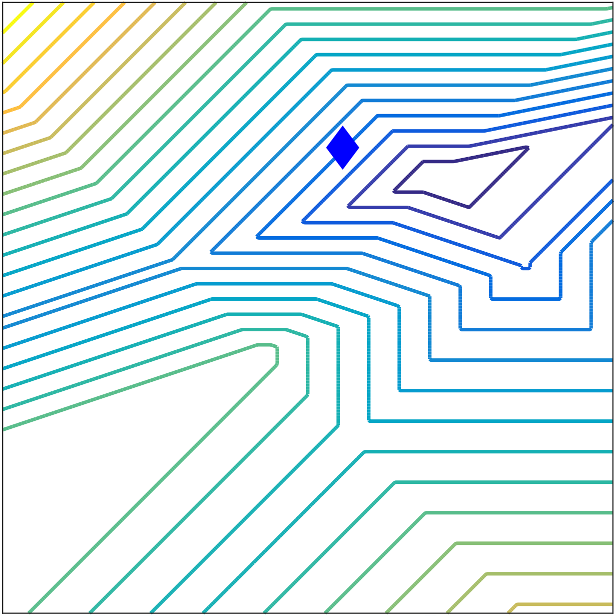
Manifold Sampling



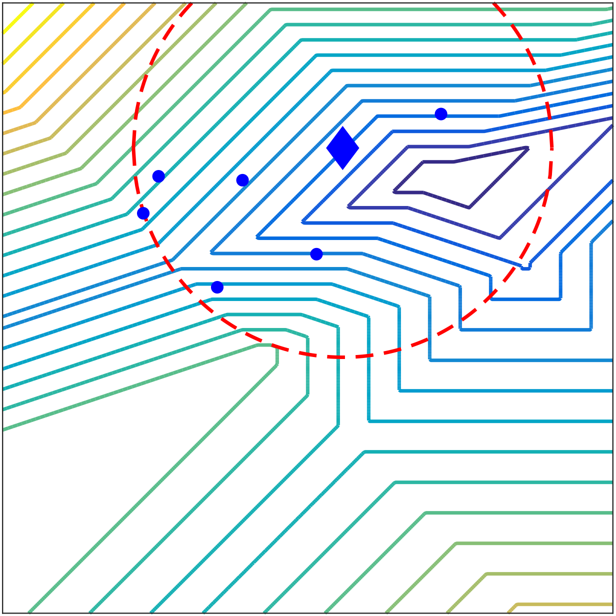
Manifold Sampling



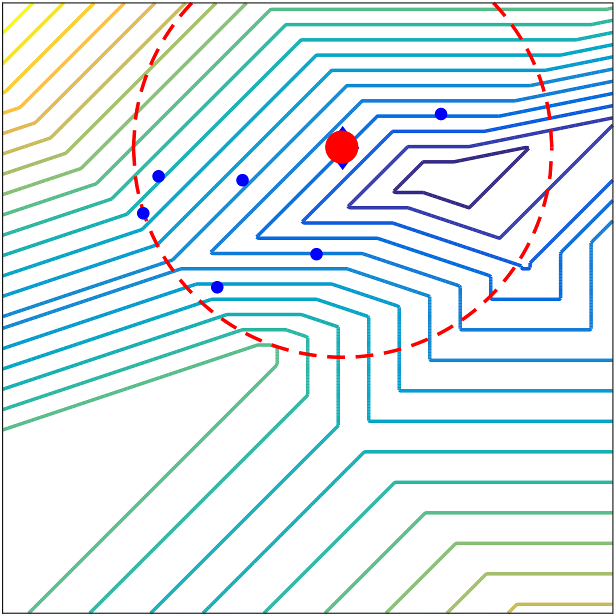
Manifold Sampling



Manifold Sampling



Manifold Sampling



Better trust-region subproblem?

Instead of solving

$$\underset{s}{\text{minimize}}\ m_k^f(x^k + s)$$
$$\textbf{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

How about

$$\underset{s}{\text{minimize}}\ h(M(x^k + s))$$
$$\textbf{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

Better trust-region subproblem?

Instead of solving

$$\begin{aligned} \underset{s}{\text{minimize}}\ & m_k^f(x^k + s) \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

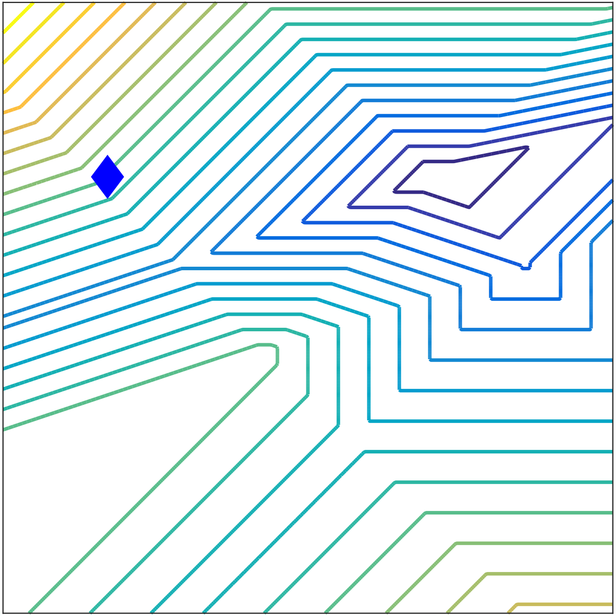
How about

$$\begin{aligned} \underset{s}{\text{minimize}}\ & h(M(x^k + s)) \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

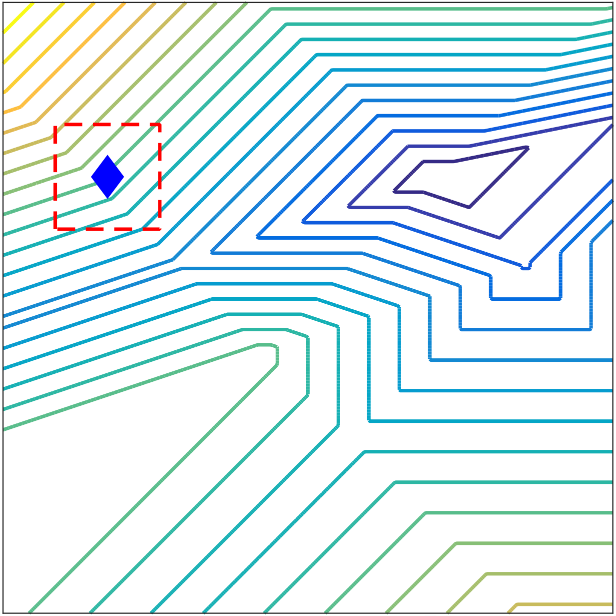
For censored ℓ_1 loss:

$$\begin{aligned} \underset{s}{\text{minimize}}\ & \sum_{i=1}^{p} |d_i - \max\{c_i, q_i(x)\}| \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

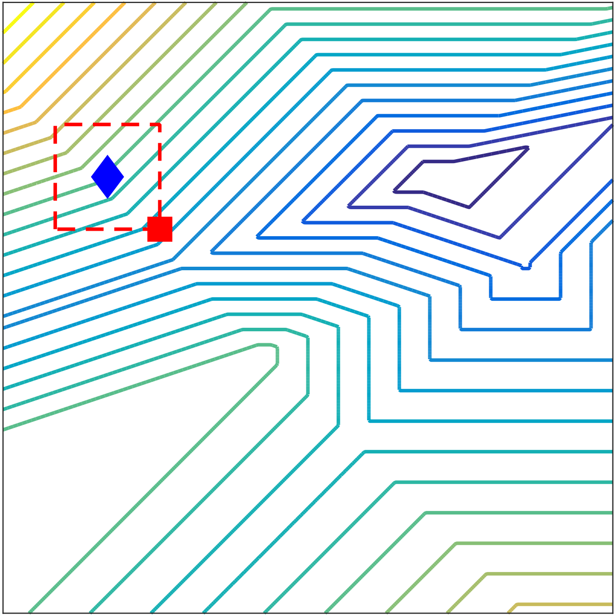
Manifold Sampling



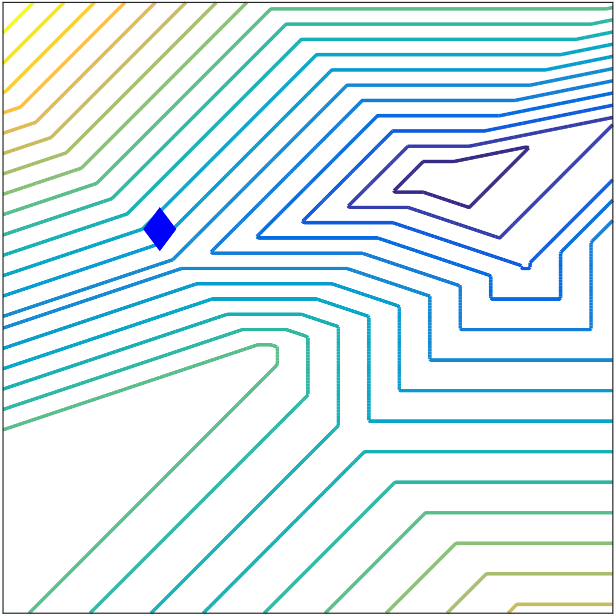
Manifold Sampling



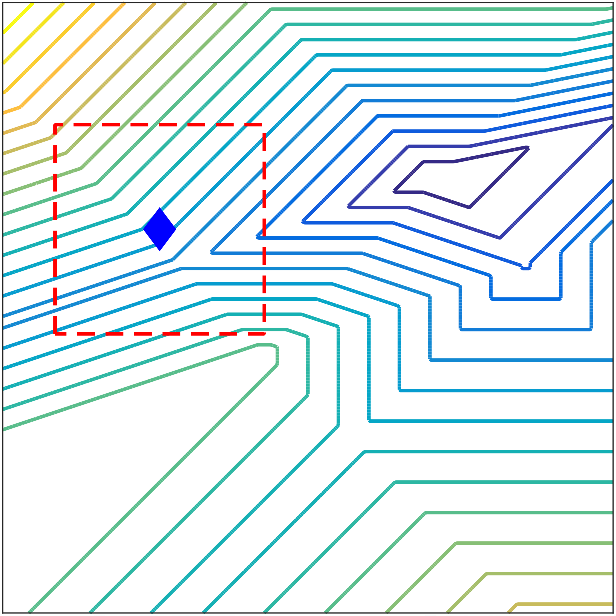
Manifold Sampling



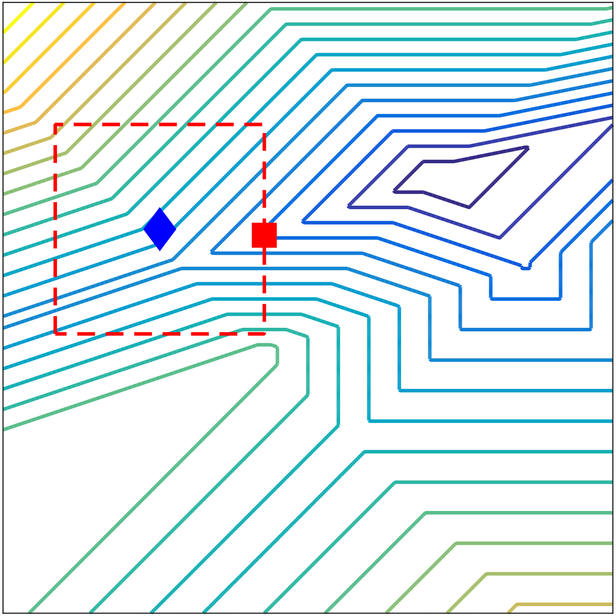
Manifold Sampling



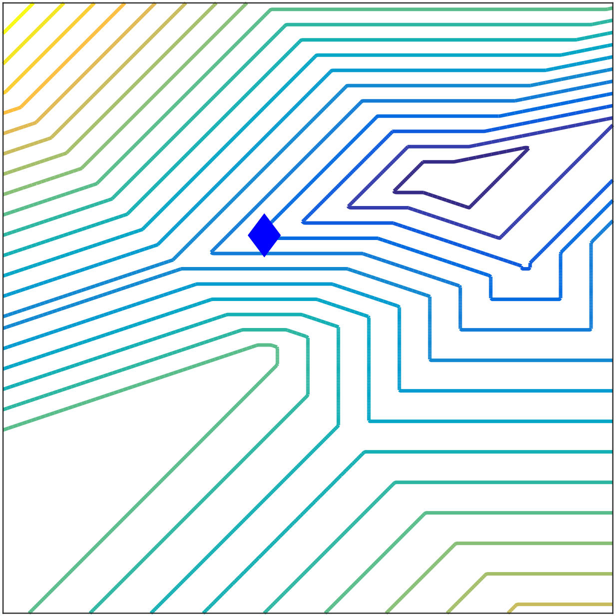
Manifold Sampling



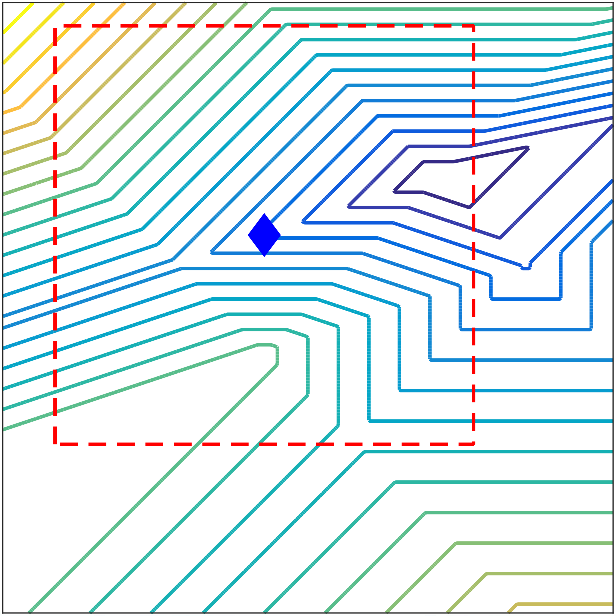
Manifold Sampling



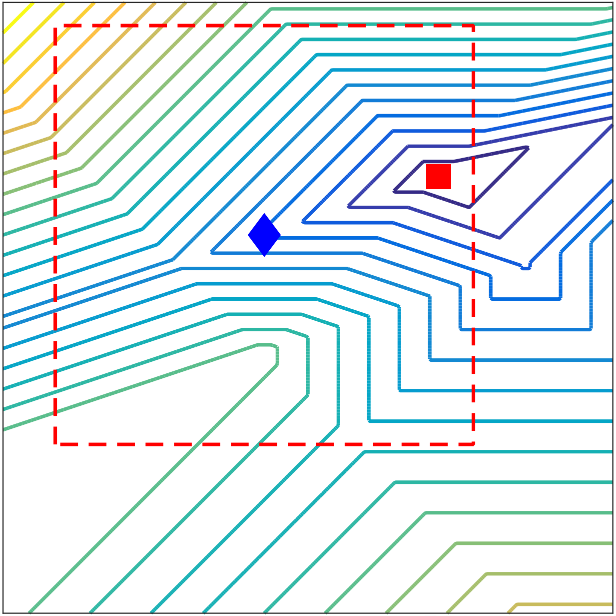
Manifold Sampling



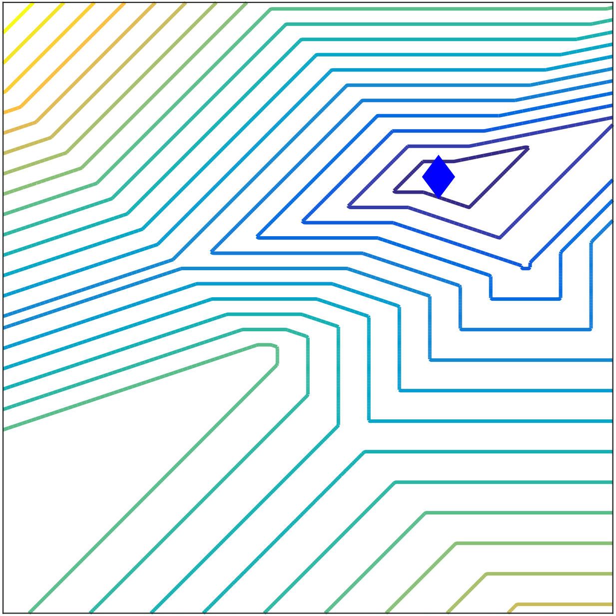
Manifold Sampling



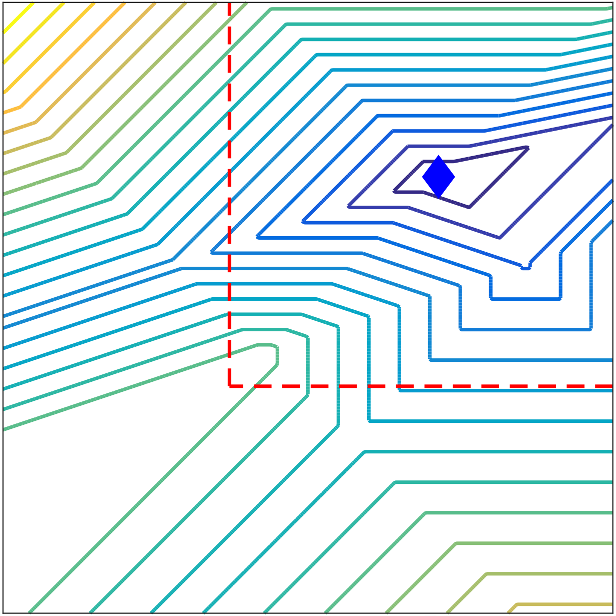
Manifold Sampling



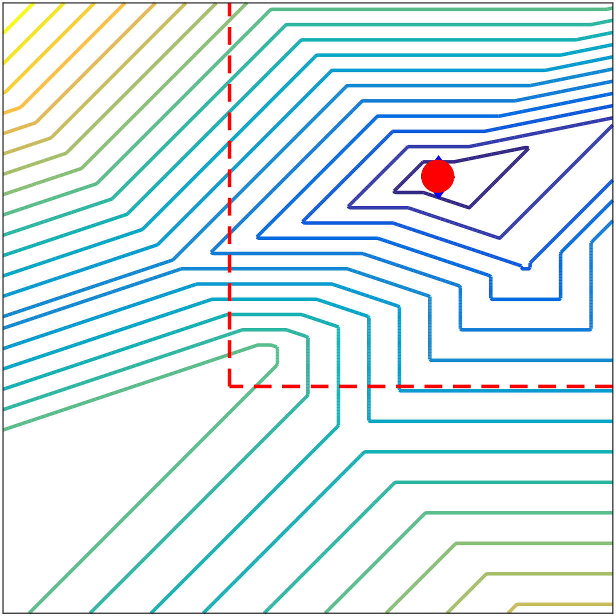
Manifold Sampling



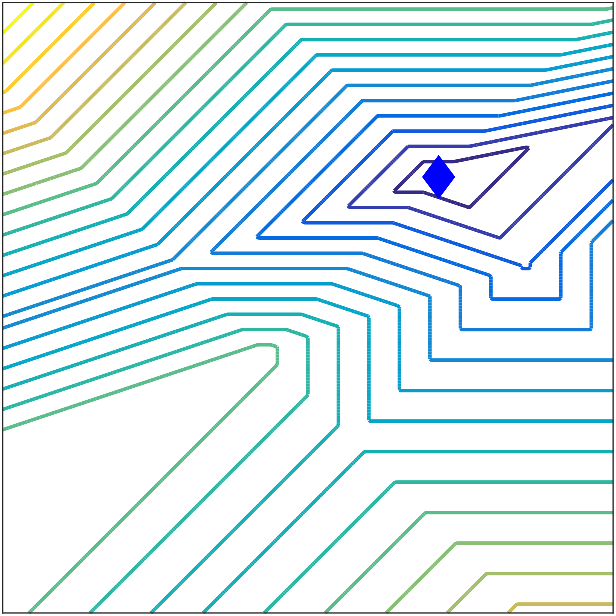
Manifold Sampling



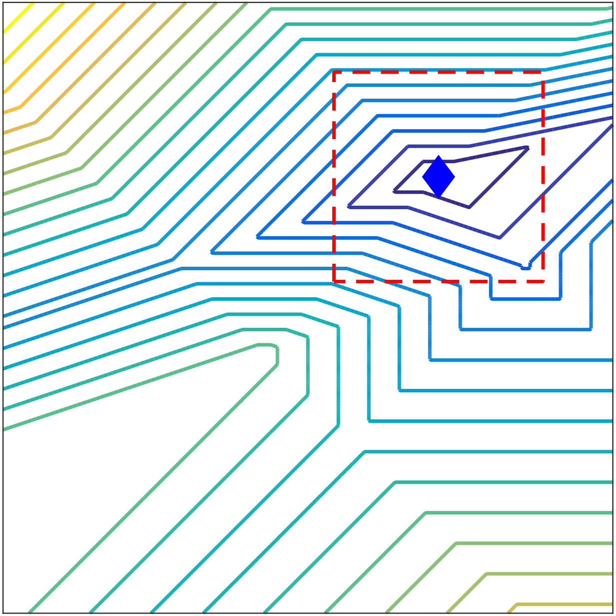
Manifold Sampling



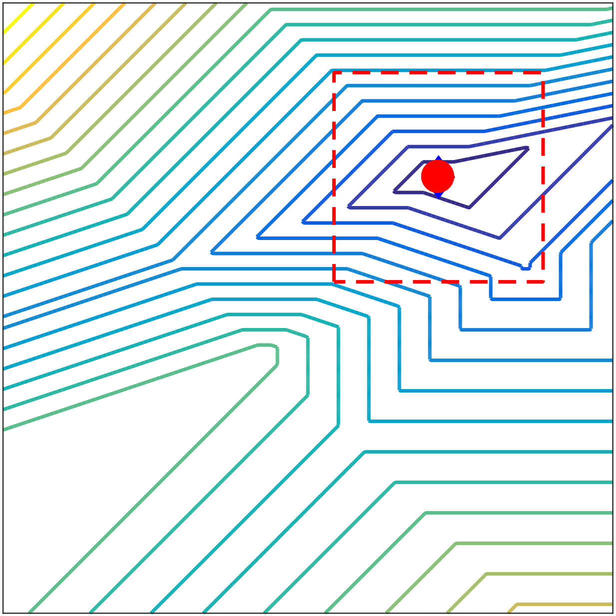
Manifold Sampling



Manifold Sampling



Manifold Sampling



Better trust-region subproblem?

Instead of solving

$$\underset{s}{\text{minimize}}\; m_k^f(x^k + s)$$
$$\text{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

How about

$$\underset{s}{\text{minimize}}\; h(M(x^k + s))$$
$$\text{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

For censored ℓ_1 loss:

$$\underset{s}{\text{minimize}}\; \sum_{i=1}^{p} \left| d_i - \max\left\{c_i, q_i(x)\right\}\right|$$
$$\text{subject to: } s \in \mathcal{B}(0, \Delta_k)$$

Better trust-region subproblem?

Instead of solving

$$\begin{aligned} \underset{s}{\text{minimize}}\ & m_k^f(x^k + s) \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

How about

$$\begin{aligned} \underset{s}{\text{minimize}}\ & h(M(x^k + s)) \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

For censored ℓ_1 loss:

$$\begin{aligned} \underset{s}{\text{minimize}}\ & \sum_{i=1}^{p} \left| d_i - \max\left\{c_i, q_i(x)\right\} \right| \\ \textbf{subject to: } & s \in \mathcal{B}(0, \Delta_k) \end{aligned}$$

Question

Best method for solving composite nonsmooth quadratic problems?

Thanks

Questions?

`jmlarson@anl.gov`